



TITLE:

2つの節をもつ正規基本形式体系の 正例からの学習 (計算機科学基礎理 論の新展開)

AUTHOR(S):

植村, 仁; 道正田, 昌; 佐藤, 優子

CITATION:

植村, 仁 ...[et al]. 2つの節をもつ正規基本形式体系の正例からの学習 (計算機科学基礎理論の新展開). 数理解析研究所講究録 2003, 1325: 116-121

ISSUE DATE:

2003-05

URL:

<http://hdl.handle.net/2433/43183>

RIGHT:

2つの節をもつ正規基本形式体系の正例からの学習

大阪府立大学大学院理学系研究科* 植村 仁 (Jin Uemura), 道正田 昌 (Masashi Dosyouda)

大阪府立大学総合科学部数理・情報科学科** 佐藤 優子 (Masako Sato)

* Graduate School of Science, Osaka Prefecture University

** Department of Mathematics and Information Sciences

Colledge of Integrated Arts and Sciences, Osaka Prefecture University

1 序

パターンとは、定数及び変数からなる有限文字列あり、パターン言語は、各変数へ定数文字列を代入して得られる文字列からなる。空列 ε 代入を許す場合、erasing パターン言語といい、許さない場合、nonerasing パターン言語という。nonerasing パターン言語の族は、Gold[6] の枠組みで正例から推論可能な言語族として、Angluin[2] によって導入された。一方、nonerasing パターン言語の族は、Shinohara[11] により extended パターン言語の名前で導入された。この言語族の推論可能性の問題は、今日まで未解決である。昨年、Reidenbach [5] によって、定数記号が 2 個の場合にこの問題を否定的に解決した。一方、パターンに含まれる変数が異なる場合、正規パターンと呼ばれる。Shinohara[11] は、erasing, nonerasing の双方に対して、正例から多項式時間で推論可能であることを示した。また、Arimura[4] は、高々 k 個のパターン言語の和の族に関して、言語の包含関係とパターン集合の構文的包摂関係の等価性を与える Compactness の概念を導入し、Compactness を有する言語和の族を効率的に学習するアルゴリズムを一般的な枠組みで構築した。そして、著者等は、Compactness が成立する必要十分条件を nonerasing(Sato[10]) と erasing(Uemura and Sato [15]) の双方に対して与えた。

本稿では、2つの節をもつ正規基本形式体系(以下、EFS)の正例からの学習問題を扱う。基本形式体系は、Smullyan[14] によって、帰納的関数の理論を再構築するために考案され、その後、Arikawa et al.[3] によって、帰納推論のための統一的な枠組みとして発展された。EFS は、論理プログラ

ミングのように機能し、形式言語を定義する機構として用いることができる。EFS は、公理(確定節)の有限集合と MP 推論による演繹体系であり、従来の EFS では、項(パターン)に含まれる変数への空列代入は、許されていなかった。本稿では、変数への空列代入を許し、さらに正規 EFS と呼ばれる節の構文に制限を課した EFS を扱う。正規 EFS は、Chomsky 階層における文脈自由言語の族を記述する能力がある(Arikawa et al.[3])。空列代入を許す場合も同じ結果が成り立つ。

2つの節が共に本体(body)を含まない場合、正規 EFS から定義される言語は、2つの erasing パターン言語の和で表され、正規言語である。一方、1つの節が本体を含む場合は、定義される言語は必ずしも正規ではないが、erasing 正規パターン言語の無限和として表すことができる。

本稿では、2つの節からなる正規 EFS を正例から効率的に学習するアルゴリズムの構築のため、erasing 正規 EFS 言語の性質、特に包含問題について考察する。正例から推論可能となる十分条件として、Wright[16], [8] は、有限の弾力性の概念を導入した。Shinohara[12] による長さ限定 EFS 言語の族をはじめ、これまでに推論可能となる具体的な言語族は、殆どこの性質を有する。残念ながら、本稿で扱う erasing 正規 EFS 言語の族は、この性質を持たないことが示される。しかし、言語族が M -有限の厚さと呼ばれる族の性質を持ち、各言語の有限証拠集合の存在が成り立つならば、正例から推論可能である([7], [9])。本稿では、 M -有限の厚さを示し、さらに、各 erasing 正規 EFS 言語の有限証拠集合を具体的に提示し、正例からの推論可能性を示している。

2 erasing 正規 EFS 言語

ここでは、正規基本形式体系 (EFS) 及び erasing 正規 EFS 言語を導入する。構文に制約がない一般的な EFS の定式化及び Chomsky 階層との対応に関しては、Arikawa[3] に詳しい。

Σ, Π を互いに交わらない有限集合とする。 Σ は通常のアルフベットである。 Π の要素は述語記号 (predicate symbol) とよばれ、それぞれの述語記号には引数 (arity) とよばれる自然数が対応付けられている。ここでは、引数が 1 である述語記号のみを扱う。さらに、 Σ, Π と交わらない可算無限集合 X を仮定する。 X の要素を変数と呼ぶ。 X の要素は x, y, z およびそれらに添字をつけた $x_1, y_1, z_1, x_2, \dots$ で表すことにする。

$(\Sigma \cup X)^*$ の要素を項 (term) またはパターン (pattern) とよび、 $\pi, \tau, \pi_1, \tau_1, \pi_2, \dots$ で表す。空列 ε もパターンであることに注意する。パターン π に含まれる変数が異なっているとき、正規パターンという。正規パターンの全体集合を \mathcal{RP} で表す。パターン π に含まれる変数の集合を V_π で表す。

$p, q_1, \dots, q_n \in \Pi$ と $\pi \in \mathcal{RP}$ に対して、次のような式を確定節 (節) という。

(i) $p(\pi) \leftarrow (ii) p(\pi) \leftarrow q_1(x_1), \dots, q_n(x_n), n \geq 0$

ただし、(ii) では、 $V_\pi = \{x_1, x_2, \dots, x_n\}$ とする。 $p(\pi)$ を頭部 (head)、 $q_1(x_1), q_2(x_2), \dots, q_n(x_n)$ を本体 (body) という。また、(i) の式を base step といい、(ii) の式で $n \geq 1$ のとき、induction step という。(ii) において、 $\pi = x$ となる場合は考えない。

定義 2.1. 正規基本形式体系 (EFS) は三項組 (Σ, Π, Γ) によって定義される。 Γ は確定節の有限集合であり、公理とよばれる。

定義 2.2. 互いに異なる x_i ($1 \leq i \leq n$) と任意のパターン $\tau_i \in \mathcal{RP}$ に対し、 $\theta = \{x_1 := \tau_1, \dots, x_n := \tau_n\}$ の形をした有限集合 (空でもよい) を代入とよぶ。 τ_i として、空列も可能である。従って、空列代入を許さない通常の基本形式体系とは異なり、代入された後の項が元の項より長さが短くなるという可能性を考えなければならない。

パターン π と代入 $\theta = \{x_1 := \tau_1, \dots, x_n := \tau_n\}$ に対して、 π 中に現れるすべての変数 x_i に τ_i を代入して得られるパターンを $\pi\theta$ で表す。

$p \in \Pi$ に対して、 $p(\pi)\theta = p(\pi\theta)$ と定義し、節 $(p(\pi) \leftarrow q_1(x_1), \dots, q_n(x_n))\theta = p(\pi)\theta \leftarrow q_1(x_1)\theta, \dots, q_n(x_n)\theta$ と定義する。

定義 2.3. 節の有限集合 Γ と集合 $\{p(w) \mid w \in \Sigma^*\}$ ($p \in \Pi$) および自然数 k に対して、関係 $\Gamma \vdash_k p(w)$ を次のように帰納的に定義する。

1. Γ 中の base step $p(\pi) \leftarrow$ と任意の代入 θ に対して、 $w = \pi\theta \in \Sigma^*$ ならば、 $\Gamma \vdash_1 p(w)$
2. $p(\pi) \leftarrow q_1(x_1), \dots, q_n(x_n) \in \Gamma$ と任意の代入 θ に対して、 $\Gamma \vdash_{k_1} x_i\theta (= w_i) (i = 1, \dots, n)$ かつ $w = \pi\theta$ ならば、 $\Gamma \vdash_k p(w)$ かつ $k = k_1 + \dots + k_n + 1$

$\Gamma \vdash_k p(w)$ が成り立つとき、 $p(w)$ は Γ から高々 k steps で証明可能であるといい、 $\Gamma \vdash_k p(w)$ なる k が存在するとき、単に $\Gamma \vdash p(w)$ と表し、 $p(w)$ は Γ から証明可能であるという。

定義 2.4. EFS $S = (\Sigma, \Pi, \Gamma)$, $p \in \Pi$ に対して、

$$L(S, p) = \{w \in \Sigma^* \mid \Gamma \vdash p(w)\}$$

と定義する。 $L(S, p)$ は Σ 上の言語である。 $L = L(S, p)$ なる EFS S と述語 p が存在するとき、言語 $L \subseteq \Sigma^*$ は erasing 正規 EFS によって定義可能である、あるいは erasing 正規 EFS 言語という。

k 個の節をもつ正規 EFS の全体集合を \mathcal{RE}^k で表し、

$$\mathcal{REL}^k = \{L(S, p) \mid S \in \mathcal{RE}^k\}$$

Arikawa et al.[3] は、通常正規 EFS 言語族が文脈自由言語族に対応することを示した。本稿で導入した erasing 正規 EFS 言語族でも同じ結果が示される。

定理 2.1. 言語 $L \subseteq \Sigma^*$ が文脈自由言語であるための必要十分条件は $L = L(S, p)$ を満たす正規 EFS $S = (\Sigma, \Pi, \Gamma)$ が存在することである。

例 2.1. 正規 EFS $S = (\{a, b\}, \{p\}, \Gamma)$ を次のように定義する。

$$p(\varepsilon) \leftarrow, \quad p(axb) \leftarrow p(x)$$

このとき、 $L(S, p) = \{a^n b^n \mid n \geq 0\}$ である。

本稿では、2つの節をもつ正規EFSの族 \mathcal{RE}^2 を扱う。節が2つしか含まないerasingEFS言語では、言語を定義する述語記号 p だけで十分であり、他の述語記号を含む節は、冗長な節であることが容易に示される。従って以下、正規EFSを単に、 Γ で表し、述語記号は p を用いる。さらに、 Γ で定義されるerasing正規言語を単に、 $L(\Gamma)$ と表す。 p が固定されているので、 Γ は次のいずれかの形で表すことができる。

- (i) $\Gamma = \{p(\pi) \leftarrow, p(\tau) \leftarrow\}$
- (ii) $\Gamma = \{p(\pi) \leftarrow, p(\tau) \leftarrow \pi(x_1), \dots, \pi(x_n)\}$

(i) の場合、 $\Gamma = \{\pi, \tau\}$ とかき、(ii) の場合は、順序対 $\Gamma = (\pi, \tau)$ で表す。(i) の場合、

$$L(\Gamma) = L(\pi) \cup L(\tau)$$

であることが容易に示される。ただし、 $L(\pi) = \{w \in \Sigma^* \mid \exists \theta \text{ s.t. } w = \pi\theta\}$ とする。 $L(\pi)$ は正規パターン言語と呼ばれ、Shinohara[11] によって、導入された。パターンを $\pi = w_1X_1w_2X_2 \dots w_mX_mw_{m+1}$ とおく。ただし、 w_1, w_{m+1} は空列も可であるが、 w_2, \dots, w_m は空列ではない。 X_i は、長さ1以上の変数列である。erasing正規パターン言語では、 ε 代入が許されているので、次の結果が示されている ([11]):

$$\pi' = w_1x_1w_2x_2 \dots w_mx_mx_{m+1}, \quad L(\pi) = L(\pi')$$

π' を標準形と呼ぶ。すべての変数に ε を代入した定数列を $c(\pi)$ で表すと、 $c(\pi) (= c(\pi')) = w_1w_2 \dots w_{m+1}$ である。 $c(\pi)$ は、 $L(\pi)$ の最短の定数列である。(i) の場合、 $w \in L(\Gamma)$ ならば、 $|w| \geq \min\{|c(\pi)|, |c(\tau)|\}$ が成り立つ。ただし、 $|\pi|$ は、 π の長さを表す。

一方、(ii) の induction step を含む EFS では、次の例で示されるように τ に対して、標準形を適用することができない。

例 2.2. $\Gamma = (aa, bx_1x_2b)$, すなわち、 $\Gamma = \{p(aa) \leftarrow, p(bx_1x_2a) \leftarrow p(x_1), p(x_2)\}$ を考える。induction step 使って得られる最短の文字列は、 $b(aa)(aa)b$ である。しかし、 bx_1x_2b の標準形 bxb に対して、 $\Gamma' = (aa, bxb)$ とすると、 $b(aa)b$ が得られる。従って、 $L(\Gamma) \neq L(\Gamma')$ である。

$\Gamma = (\pi, \tau)$ において、 τ のすべての変数に、パターン π を代入して得られる正規パターンを τ_π

と表すと、induction step を一度以上使って得られる定数列の最短の長さは、 $|c(\tau_\pi)|$ である。故に、 $w \in L(\Gamma)$ ならば、 $|w| \geq \min\{|c(\pi)|, |c(\tau_\pi)|\}$ が成り立つ。

定義 2.5. EFS Γ は、任意の $\Gamma' \subset \Gamma$ に対して、 $L(\Gamma) \subsetneq L(\Gamma')$ のとき、既約 (reduced) という。

3 正例からの帰納推論

文字列の無限列 w_1, w_2, \dots が言語 L の正提示であるとは、 $\{w_n \mid n \geq 1\} = L$ が成り立つことである。文字列の無限列 $\sigma = w_1, w_2, \dots$ の n 番目までの初期部分列を $\sigma[n]$ で表す。

言語族 $\mathcal{L} = L_1, L_2, \dots$ が帰納的言語の添字付き族 (indexed family of recursive languages) であるとは、次のような計算可能関数 $f: N \times \Sigma^* \rightarrow \{0, 1\}$ が存在することをいう: $f(i, w) = 1$, if $w \in L_i$; 0 , if $w \notin L_i$. ここでの添字 i は言語を定義するオートマトンや形式文法等を意味すると考えられるが、本稿では前節で扱った正規EFS等を意味する。

推論機械 M とは、次々に入力を要求し、次々に出力を生成する実行的な手続きのことであり、 M が生成する出力を推測と呼ぶ。文字列の無限列 σ に対して、その有限列 $\sigma[n]$ が入力された後、 M が生成する推測を $M(\sigma[n])$ で表す。推論機械 M が入力の列 σ に対して、添字 $g \in N$ に収束するとは、ある $m \in N$ が存在し、任意の $n \geq m$ に対して、 $M(\sigma[n]) = g$ となることをいう。推論機械 M が言語 L を正例から極限同定 (identification in the limit) する (または、単に推論する) とは、 L の任意の正提示に対して、 M が $L_i = L$ となる添字 i に収束することをいう。また、任意の言語 $L \in \mathcal{L}$ を正例から極限同定する推論機械 M が存在するとき、言語族 \mathcal{L} は正例から推論可能であるという。

推論機械 M が言語族 \mathcal{L} を正例から推論し、各入力を受け取ってから推測を出力するまでに必要な時間がそれまでの入力の長さの和のある多項式で押えられる時、その推論機械 M は言語族 \mathcal{L} を正例から多項式時間 (の更新) で推論するという。また、言語族 \mathcal{L} を正例から多項式時間 (の更新) で推論する推論機械が存在する時、その言語族は、正例から多項式時間 (の更新) で推論可能であると

いう。

Angluin[2] は、正例からの推論可能性の特徴づけ定理を証明し、更に、有限の厚さ等推論可能性の十分条件を提示した。一方、Wright[16], [8] は、有限の厚さを一般化した有限の弾力性の概念を導入した。言語族 \mathcal{L} が有限の弾力性を持つとは、次の集合的性質を満たす定数列の無限列 w_0, w_1, \dots 、及び \mathcal{L} の言語の無限列 L_1, L_2, \dots が存在しないことをいう。任意の $k \geq 1$ に対して、 $\{w_0, w_1, \dots, w_k\} \subseteq L_k$, $w_{k+1} \notin L_k$ である。

有限の弾力性は、(i) 正例からの推論可能であるための十分条件であり、(ii) 言語族の和 ([16])・積・連接等 ([7], [9]) の演算に関して閉じている等、極めて良い性質を有している。Shinohara[12] による長さ限定 EFS を始めこれまでに正例から推論可能な具体的な言語族は、殆どこの性質を有している。

定理 3.1. *erasing* 正規言語族 \mathcal{REL}^2 は有限の弾力性をもたない。

一方、Moriyama & Sato[7] は、 M -有限の厚さという有限の厚さを一般化した概念を導入した。空でない有限集合 $S \subseteq \Sigma^*$ に対して、

$$\text{MIN}(S, \mathcal{L}) = \{L \in \mathcal{L} \mid L \text{ は } S \text{ の極小言語}\}$$

定義 3.1. 言語族 \mathcal{L} が次の条件を満たすとき、 M -有限の厚さをもつという。任意の有限集合 $S \subseteq \Sigma^*$ に対して、(i) $\#\text{MIN}(S, \mathcal{L}) < \infty$, (ii) $S \subseteq L, L \in \mathcal{L}$ ならば、 $L' \subseteq L$ となる $L' \in \text{MIN}(S, \mathcal{L})$ が存在することである。

M -有限の厚さは、正例から推論可能であるための十分条件ではないが、有限の弾力性と同様、様々な族演算に関して閉じている ([7], [9])。更に、 M -有限の厚さをもつ族では、正例からの推論可能性が「各言語が、有限証拠集合 (finite tell-tale) をもつこと」という集合論的な性質に帰着できる ([9])。ここで、言語 $L \in \mathcal{L}$ の有限証拠集合 $S \subseteq \Sigma^*$ とは、 $L \in \text{MIN}(S, \mathcal{L})$ を満たす有限集合をいう ([2])。

定理 3.2. 言語族 $\mathcal{L} = L(\Gamma_1), L(\Gamma_2), \dots$ が M -有限の厚さを持ち、各言語 $L(\Gamma) \in \mathcal{L}$ の有限証拠集合が存在するならば、正例から推論可能である。

4 2つの節からなる正則基本形式の学習

本節では、1つの basic step 及び 1つの induction step からなる正規 EFS $\Gamma = (\pi, \tau)$ について、特に $\pi \notin \Sigma^*$ を考える。

V_π, V_τ を π, τ に含まれる変数とし、 $V_\tau = \{x_1, x_2, \dots, x_n\}$ とする。 $V_\tau \neq \emptyset$ であることに注意する。パターン τ のすべての変数に π を代入したパターンを τ_π で表し、さらに $\Gamma_\tau = \bigcup_{t=1}^{\infty} \Gamma_t$ を導入する。ただし、 Γ_t は、次のように帰納的に定義する：

$$\begin{aligned} \Gamma_1 &= \{\tau_\pi\}, \text{ where } \tau_\pi = \tau\{x_i := \pi(i=1, \dots, n)\}, \\ \Gamma_t &= \Gamma_{t-1} \cup \{\tau\{x_1 := \xi_1, x_2 := \xi_2, \dots, x_n := \xi_n\} \\ &\quad \mid \xi_i \in \Gamma_{t-1} \cup \{\pi\} (i=1, 2, \dots, n)\} \quad (t \geq 2) \end{aligned}$$

ただし、代入される変数名は、すべて異なるものとする。明らかに、任意の $\xi \in \Gamma_\tau$ に対して、 $\xi \preceq \tau$ である。

補題 4.1. $\text{EFS} = (\pi, \tau), V_\tau = \{x_1, \dots, x_n\}$ とし、 $\xi_1, \xi_2, \dots, \xi_n$ をパターンとする。 $L(\xi_i) \subseteq L(\Gamma) (i=1, \dots, n)$ ならば、 $L(\theta) \subseteq L(\Gamma)$ である。ただし、 $\theta = \{x_1 := \xi_1, \dots, x_n := \xi_n\}$ とする。

定理 4.1. $\text{EFS} = (\pi, \tau)$ に対して、

$$L(\Gamma) = L(\pi) \cup L(\Gamma_\tau).$$

上記の補題より、induction step を含む EFS 言語は、パターン言語の和をして表現できることがわかる。しかし、この和は、無限個の和である。さらに、次のパターンを定義する。

$$\bar{\tau}_\pi = \tau\{y := x\pi x', y \in V_\tau\} (= \tau_{x\pi x'})$$

次の結果は、 $\bar{\tau}_\pi, \Gamma_\tau$ の定義から導ける。

補題 4.2. $\Gamma = (\pi, \tau)$ とする。このとき、(1) 任意の $\xi \in \Gamma_\tau$ に対して、 $\xi \preceq \bar{\tau}_\pi$, (2) $L(\Gamma_\tau) \subseteq L(\bar{\tau}_\pi)$

パターン言語 (の和) の包含問題では、パターンがどのような文字で始まり、どのような文字で終わっているかの情報が重要な key である。パターン π の接頭語で最長の定数列を π の接頭定数列 (ε も可能) といい、 A_π で表す。 π の接尾定数列

も同様に定義し, B_π で表す. $\pi = w(x\pi'x')w'$ ならば, $A_\pi = w, B_\pi = w'$ である. また, 定数列 $\pi \in \Sigma^*$ ならば, $A_\pi = B_\pi = \pi$ であり, π が変数を含むならば, $\pi = A_\pi(x\pi'x')B_\pi$ ($\pi' = \varepsilon$ も可) と表すことが出来る.

Σ^* 上の関係 \preceq_p, \preceq_s を次のように定義する: $v \in \Sigma^*$ が $w \in \Sigma^*$ の接頭語であるとき, $v \preceq_p w$ で表す. また, v が w の接尾語の場合, $v \preceq_s w$ で表す. 明らかに, \preceq_p, \preceq_s は共に, 半順序関係である.

補題 4.3. $\text{EFS}\Gamma = (\pi, \tau), \xi \in \Gamma_\tau$ とする. このとき, $A_\xi = A_\tau^{i_0}A_\pi$, $B_\xi = B_\pi B_\tau^{j_0}$ を満たす $i_0, j_0 \geq 1$ が存在する.

文字列の対 (w, v) は, $w \preceq_p v$ または, $v \preceq_p w$ であるとき, prefix の関係があるといい, $(w, v) \in \text{Pref}$ で表す. 同様に, Suff も関係 \preceq_s に対して定義する.

補題 4.4. 変数を含むパターン π, τ に対して, 次が成り立つ. $L(\pi) \cap L(\tau) \neq \emptyset \iff (A_\pi, A_\tau) \in \text{Pref}, (B_\pi, B_\tau) \in \text{Suff}$

次の結果は, 明らかである.

補題 4.5. π, τ を変数を含むパターンとする. $\pi = A_\pi(x\pi'x')B_\pi$ 及び, $\tau = A_\tau(x\tau'x')B_\tau$ に対して, τ' は部分列として, π' を含むとする. このとき, $\tau \preceq \pi \iff A_\pi \preceq_p A_\tau, B_\pi \preceq_s B_\tau$

補題 4.6. $\Gamma = (\pi, \tau)$ とする. $A_\pi = B_\pi = \varepsilon$ ならば, $L(\Gamma_\tau) \subseteq L(\pi)$, すなわち, $L(\Gamma) = L(\pi)$.

4.1 既約 EFS の特徴付け

$\Gamma = (\pi, \tau)$ が既約となるのは, $L(\pi) \subsetneq L(\Gamma)$, すなわち, $L(\Gamma_\tau) \not\subseteq L(\pi)$ のときである. 本節では, 既約 EFS の性質について論じる.

補題 4.6 より, π が変数で始まり, 変数で終わるパターン (v - v パターンという) ならば, $\Gamma = (\pi, \tau)$ は既約ではない. また, $\pi \in \Sigma^*$ ならば, $\Gamma = (\pi, \tau)$ は既約である. 以下, EFS が既約となる条件について考察する. 文字列 $w \in \Sigma^*$ は, $w \preceq_p v^i$ となる $i \geq 0$ が存在するとき, $w \preceq_p v^*$ で表す. $w \preceq_s v^*$ も同様に定義する.

補題 4.7. $w \in \Sigma^*, v \in \Sigma^+$ とする.

$$(i) \exists i_0 \geq 1 \text{ s.t. } w \preceq_p v^{i_0}w \iff w \preceq_p v^* \\ \iff \forall i \geq 0, w \preceq_p v^i w$$

$$(ii) \exists j_0 \geq 1 \text{ s.t. } w \preceq_s wv^{j_0} \iff w \preceq_s v^* \\ \iff \forall j \geq 0, w \preceq_s wv^j$$

定理 4.2. $\text{EFS } \Gamma = (\pi, \tau)$ に対して, 次の命題は等価である. (1) Γ は既約である. (2) $A_\pi \not\preceq_p A_\tau^* (A_\tau \neq \varepsilon)$ または, $B_\pi \not\preceq_s B_\tau^* (B_\tau \neq \varepsilon)$ (3) $L(\Gamma_\tau) \cap L(\pi) = \emptyset$

上記の定理より, 次の結果が得られる.

系 4.1. 与えられた $\text{EFS } \Gamma = (\pi, \tau)$ が既約か否かの決定問題は, $O(|\pi| + |\tau|)$ の時間で計算可能である.

系 4.2. π と τ のいずれかが v - v パターンならば, $\text{EFS } \Gamma = (\pi, \tau)$ は既約ではない.

補題 4.8. $\text{EFS } \Gamma = (\pi, \tau)$ は既約とし, γ をパターンとする. $L(\gamma) \subseteq L(\Gamma)$ のとき, $\gamma \not\preceq \pi \iff L(\gamma) \subseteq L(\Gamma_\tau)$

ここで, $\Gamma_{\tau*} = \{\xi \in \Gamma_\tau \mid L(\tau_\pi) \cap L(\xi) = \emptyset\}$ と定義しておく.

Γ が既約であるとき, 補題 4.4 より, $\xi \in \Gamma_{\tau*}$ ならば, $(A_\xi, A_{\tau_\pi}) \notin \text{Pref}$ または $(B_\xi, B_{\tau_\pi}) \notin \text{Suff}$ が成立する.

補題 4.9. 既約な $\text{EFS } \Gamma = (\pi, \tau)$ に対して, $\xi \in \Gamma_\tau, L(\xi) \cap L(\tau_\pi) \neq \emptyset \Rightarrow \xi \preceq \tau_\pi$

系 4.3. Γ が既約であるとき, $L(\Gamma_\tau) = L(\tau_\pi) \cup L(\Gamma_{\tau*}), L(\tau_\pi) \cap L(\Gamma_{\tau*}) = \emptyset$

補題 4.10. $\text{EFS } \Gamma = (\pi, \tau)$ は既約とし, γ をパターンとする. $L(\gamma) \subseteq L(\Gamma_\tau)$ のとき, $\gamma \not\preceq \tau_\pi \iff L(\gamma) \subseteq L(\Gamma_{\tau*})$

パターン $\pi \in \mathcal{RP}$ の各変数に ε かまたは, 1 定数文字を代入して得られる $L(\pi)$ の有限部分集合を $S(\pi)$ で表す.

補題 4.11. $\text{EFS } \Gamma = (\pi, \tau)$ は既約とし, γ をパターンとする. $S(\gamma) \subseteq L(\Gamma_\tau)$ のとき,

$$S(\gamma) \not\subseteq L(\tau_\pi) \iff L(\gamma) \subseteq L(\Gamma_{\tau*})$$

定理 4.3. $\text{EFS } \Gamma = (\pi, \tau), \Gamma' = (\alpha, \beta)$ は既約とする. このとき, $L(\Gamma') \subseteq L(\Gamma)$ ならば, 次の等

価性が成り立つ.

- (i) $L(\Gamma') \subseteq L(\pi) \iff \alpha \preceq \pi, \beta_\alpha \preceq \pi,$
- (ii) $L(\alpha) \subseteq L(\pi), L(\Gamma'_\beta) \subseteq L(\Gamma_\tau)$
 $\iff \alpha \preceq \pi, \beta_\alpha \not\preceq \pi,$
- (iii) $L(\alpha) \subseteq L(\Gamma_\tau), L(\Gamma'_\beta) \subseteq L(\pi)$
 $\iff \alpha \not\preceq \pi, \beta_\alpha \preceq \pi, \beta_{\beta_\alpha} \preceq \pi.$

4.2 $L(\Gamma)$ の有限証拠集合

この節では, 正規 EFS 言語の族が正例から帰納推論可能であることについて考える. そのためには, この族が M -有限の厚さをもつこと, 各 EFS 言語が有限証拠集合をもつことを確認する必要がある.

定理 4.4. 正規 EFS 言語の族 \mathcal{REL}^2 は, M -有限の厚さをもつ.

EFS $\Gamma = \{\pi, \tau\}$ に対して, $L(\Gamma)$ の有限部分集合 $S(\Gamma)$ を定義する: $S(\Gamma) = S(\pi) \cup S(\tau)$. $\#\Sigma \geq 4$ のとき, これは $L(\Gamma)$ の有限証拠集合となること が示されている ([15]).

また, induction step を含む既約な EFS $\Gamma = (\pi, \tau)$ に対して, $S(\Gamma) = S(\pi) \cup S(\tau_\pi) \cup S(\tau_{\tau_\pi}) \cup S(\tau_{\tau_{\tau_\pi}})$ $L(\Gamma), L(\pi)$ の最短の定数列は, $c(\pi)$, $L(\Gamma_\tau)$ の最短の定数列は, $c(\tau_\pi)$ だけであることを注意する.

定理 4.5. 任意の既約正規 EFS $\Gamma \in \mathcal{REL}^2$ に対して, $S(\Gamma)$ は, $L(\Gamma)$ の (族 \mathcal{REL}^2 における) 有限証拠集合である

定理 4.4, 4.5 より, 次の結果が得られる.

定理 4.6. 正規 EFS 言語の族 \mathcal{REL}^2 は, 正例から帰納推論可能である.

参考文献

- [1] D. Angluin: *Finding patterns common to a set of strings*, Information and Control, **21**, 46-62, (1980).
- [2] D. Angluin: *Inductive inference of formal languages from positive data*, Information and Control, **45**, 117-135, (1980).
- [3] S. Arikawa, T. Shinohara and A. Yamamoto: *Elementary Formal Systems as a Unifying Framework for Language Learning*, Proc. 2nd COLT, 312-327, (1990).
- [4] H. Arimura, T. Shinohara and S. Otsuki: *Finding minimal generalizations for unions of pattern languages and its application to inductive inference from positive data*, Lecture Notes in Computer Science, **775**, 646-660, (1994).
- [5] D. Reidenbach: *Result on Inductive Inference of Extended Pattern Languages*, Lecture Notes in Artificial Intelligence, **2533**, 308-320, (2002).
- [6] E. M. Gold: *Language identification in the limit*, Information and Control, **10**, 447-474, (1967).
- [7] T. Moriyama and M. Sato: *Properties of language classes with finite elasticity*, IEICE Transactions on Information and Systems, **E78-D(5)**, 532-538, (1995).
- [8] T. Motoki, T. Shinohara and K. Wright: *The correct definition of finite elasticity: corrigendum to identification of unions*, in Proceedings of the Fourth Annual ACM Workshop on Computational Learning Theory, 375-373, (1991).
- [9] M. Sato: *Inductive Inference of Formal Language*, Bull. Inf. Cybern., **27-1**, 85-106, (1995).
- [10] M. Sato, Y. Mukouchi and D. Zheng: *Characteristic sets for unions of regular pattern languages and compactness*, Lecture Notes in Artificial Intelligence, **1501**, 220-233, (1998).
- [11] T. Shinohara: *Polynomial time inference of extended regular pattern languages*, RIMS Symposia on Software Science and Engineering, Kyoto, 1982, Proceedings, Lecture Notes in Computer Science, **147**, 115-127, (1982).
- [12] T. Shinohara: *Rich classes inferable from positive data: length-bounded elementary formal systems*, Information and computation, **108**, 175-186, (1994).
- [13] T. Shinohara and H. Arimura: *Inductive inference of unbounded unions of pattern languages from positive data*, Lecture Notes in Artificial Intelligence, **1160**, 256-271, (1996).
- [14] R.M. Smullyan: *Theory of Formal Systems*, Princeton University Press, Princeton, New Jersey, (1961).
- [15] J. Uemura and M. Sato: *Compactness and Learning of Unions of Erasing Regular Pattern Languages*, Lecture Notes in Artificial Intelligence, **2533**, 293-307, (2002).
- [16] K. Wright: *Identification of unions of languages drawn from positive data*, Proc. the 2nd Annual Workshop on Computational Learning Theory, 328-333, (1989).